Exploring Convolutional Neural Networks: A Study and Investigation Based on Face Recognition

Ruicheng Li*

School of Intelligent Engineering, Xi'an Jiaotong-Liverpool University, Jiangsu, 215123, China

* Corresponding Author Email: Ruicheng.Li22@student.xjtlu.edu.cn

Abstract. Machine learning (ML) is profoundly transforming decision-making processes across a range of industries, simultaneously presenting challenges to existing labor markets and facilitating significant technological and innovative advancements. This paper undertakes a thorough exploration of the broad societal implications of ML and its pivotal contributions towards achieving sustainable development goals. It provides an in-depth examination of the foundational theories underpinning machine learning and presents a detailed taxonomy of ML algorithms, with a particular emphasis on the roles of convolutional and pooling layers in deep learning models utilized for image recognition tasks. Through rigorous evaluative analyses of various models, this study offers valuable insights into the comparative efficacy of these models. Specifically, the FaceRecognition model is highlighted for its superior accuracy, reduced latency, and enhanced throughput in face recognition tasks, outperforming established benchmarks. However, the study also identifies a notable deficiency in the model's capability for object classification. Recommendations are made for reducing the complexity of models while simultaneously increasing their diversity, aiming to enhance overall performance. Furthermore, the paper underscores the significant impact these technologies on societal decision-making processes, emphasizing the need for careful consideration of both the potential benefits and the ethical implications associated with machine learning deployment.

Keywords: Machine learning, Deep learning model, Convolution layers, Pooling layers.

1. Introduction

Machine learning (ML) is a pivotal subfield of artificial intelligence (AI) that has catalyzed a paradigm shift in data-driven decision-making and analytical processes across various industries [1]. It has been instrumental in enhancing operational efficiencies and fostering innovation, while simultaneously posing challenges to labor market dynamics and necessitating the recalibration of workforce skill sets [2]. The societal implications of ML are multifaceted, encompassing ethical, privacy, and regulatory considerations that require a nuanced and collaborative approach to governance [3]. The global technological landscape is marked by an intense competition in ML, with nations strategically prioritizing AI development to gain a competitive edge in the knowledge economy [4]. Additionally, ML's contributions to sustainable development are increasingly recognized, with its applications in environmental monitoring and resource management offering potential solutions to pressing global sustainability challenges [5].

This scholarly paper commences with an exposition of the foundational theories in machine learning, thereby establishing a solid foundation for the comprehension of its cardinal principles. Subsequently, it delineates a fundamental taxonomy of machine learning algorithms, thereby erecting a framework conducive to the discernment of diverse methodologies within the discipline. The discourse then transitions into an exhaustive examination of the convolutional and pooling layers, which are pivotal components in the architecture of deep learning models, particularly pertinent to tasks of image recognition. The paper meticulously scrutinizes the roles and contributions of these layers to the overall performance of the models. Following this, the paper embarks on an evaluative and comparative analysis of multiple models, meticulously assessing their performance metrics to discern their respective merits and limitations. The synthesis of these evaluations yields profound insights into the relative efficacy of the models under scrutiny. In the concluding section, the paper succinctly summarizes the pivotal points and proposes potential avenues for future research that could propel the field's advancement.

2. Theoretical basis

2.1. Machine learning basic

In engineering and computer science, Machine Learning is about teaching computers to learn from data without being explicitly programmed [6]. It can be simply described as: using the programs to let the machine performs like humans. Among the categories of machine learning, the three most common ones are supervised learning, unsupervised learning and reinforcement learning. For the supervised learning, every element has its own label and the program can distinguish by the differences of the labels, while in the unsupervised learning system, the elements do not have labels and the programs clarify the elements into kinds of classes by whether their features and characteristics are the same or not. For instance, the spatial motion features for human action recognition, a kind of KNN-Based machine learning classifier which is used on deep learning, which can achieve the human actions recognition by evaluating the violence detection, surveillance video, and some other things [7]. And for the reinforcement learning, every route has the feedback which can be negative or positive, which means that the positive the feedback is, the better this route is and the negative the opposite. It can calibrate the parameters timely according to the environment changing to let the algorithm greater. The machine learning is widely used in several fields, Data mining and machine learning (ML) techniques offer significant potential for enhancing the endeavors of sales researchers and managers, alongside augmenting the sales process and enriching customer experiences [8]. Recent advancements in artificial intelligence (AI) have underscored the utility of machine learning (ML) for organizations aiming to extract value from data. Nevertheless, given ML's predominant association with technical domains like computer science and engineering, integrating ML training into non-technical educational curricula, such as social sciences courses, poses notable challenges [9]. Using machine learning for facial recognition is also an important research direction in the field of computer vision. The process of convolution and pooling layers is shown in figure 1.

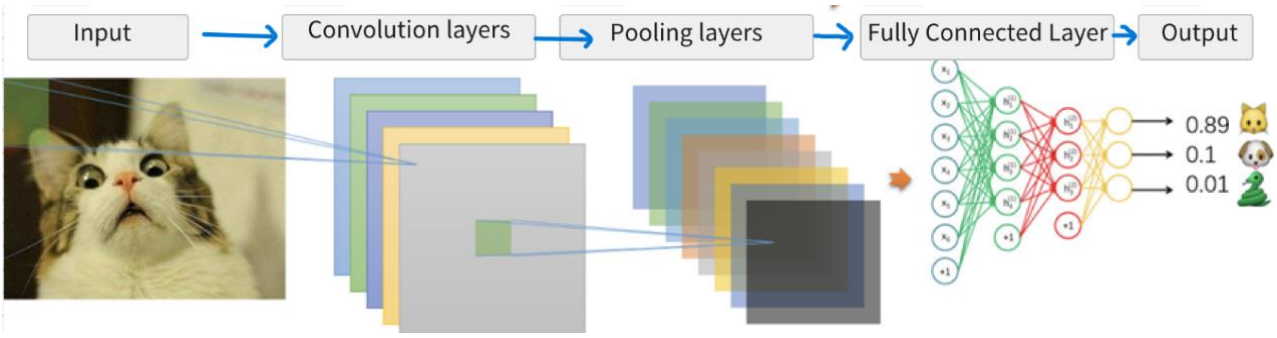

Fig. 1 The process of convolution and pooling layers (Photo/Picture credit: Original)

2.2. Convolution layer

In advanced mathematics, convolution is defined as a bilinear transformation of two functions f and g, often denoted as $f*g$, resulting in a third function. It's essentially a type of integral transform, representing the integral of the product of f and g over the overlapping range, with one function flipped and shifted. Convolution is formally expressed as:

$$\int_{-\infty}^{\infty} f(\tau)g(x-\tau)\,d\tau \tag{1}$$

where the integration denotes the convolution of functions f and g at point t, and τ represents the integration variable.

To explain convolution in CNN, f represents the relationship between τ and f, while g represents the influence that changes based on both τ and x. The integration sums up f over any period. In CNN, this logic is similar, and distinguishing between f and g is challenging.

In CNNs, images are composed of abundant pixel points, and convolutional kernels process this information. The process can be simplified as dividing pixel points into 3x3 matrices, while the

convolutional kernel is also a 3x3 matrix. Multiplying corresponding elements and summing them up gives a single convolution result. This operation integrates influences of pixel points or filters information, allowing a machine to represent pixel points in a matrix. Convolution extracts features from environmental images.

Convolutional Neural Networks (CNNs) are tailored for image recognition tasks, capitalizing on their capacity to detect hierarchical patterns within visual data. Engineered with the capability to decompose images into constituent elements, they augment classification precision, particularly in complex endeavors such as CDW sorting [10]. In contrast to traditional image or scene recognition systems, neural networks have the capability to automatically learn features. Deep neural networks, comprising millions of parameters, necessitate extensive datasets for precise feature acquisition. Leveraging the abundance of large datasets containing millions of images, convolutional networks excel in acquiring task-specific features with remarkable discriminative power. The effective utilization of deep convolutional neural networks (CNNs) in image classification has led to the widespread adoption of neural networks for various scene comprehension tasks [11]. And for face recognition, the convolution layers are needed to get the features from the images.

2.3. Pooling layer

As what the article has introduced, the convolution layers are used for extracting features, but in most common situations, the parameters which are abstracted from the environment are too many to deal with and the solution of it is to use the pooling layers to decrease the parameters. Several convolution results will be organized to one result by being deposed by the pooling window and the length of step equals to the length of pooling layers' side.

Using pooling layers in CNN can achieve lots of thoughts and solve difficulties. For instance, Utilizing Machine Learning and Deep Learning for Malaria Parasite Identification in Conventional Microscopic Blood Smear Images and global average pooling (GAP) are both extensions to the field of pooling [12, 13]. And for the face recognition, the pooling layers are important for reducing the features of the faces to let the computer system analysis easily to increase the precision of the model.

3. Code and model architecture

3.1. The preparation of data set

To train a model, it needs a training which is a set of examples used for learning, which is to fit the parameters of the classifier and a validation set to optimizing the parameters (e.g., architecture, not weights) of a classifier, such as selecting the number of hidden units in a neural network [14]. These two sets will participate in training the performance of the model and the model can determine different labels by what the training set has been put up, while the labels will be confirmed whether it is correct or not by the validation set. The system will loop this process and improve the accuracy and performance of the model. Finally, to test the versatility of the model, the testing set will be used to assess whether the model is great enough to or not, but the testing set itself will not be part of model training and it is only for testing. What the problem might appear is that if the model has a high accuracy and great performance with low loss function value on the training set and the opposite on the testing set, it shows that the model might have exhibited overfitting tendencies, where it predominantly memorized the training dataset instead of effectively discerning and generalizing underlying data features [15, 16]. To avoid this, increasing the model complexification, adding more features to enhance input data representation, adjusting parameters and hyperparameters and lowering regularization constraints are effective. And in the face recognition project, the data set is found from Kaggle open datasets, the first 80 percent is set to training set and the last 20 percent is set to testing set.

The model developed within the confines of this project, denoted as "FaceRecognition," undergoes rigorous evaluation encompassing metrics such as accuracy, latency, and throughput, thereby

enabling a comprehensive comparative analysis against benchmark models, namely "ResNet-50" and "Inception_v3."

3.2. The model performance and the comparation

The model which is created by the project only includes three and convolution layers and three pooling layers to deal with the parameters, and a flatten layer to change the data form from matrix to one-dimensional data to let the system connect. The structure is simple but the effect is powerful. The final training accuracy is about 97 percent, the final training loss is about 0.082, the final validation accuracy is about 0.49 with 6.38 of validation loss. These are only the basic performance of the model, and to test whether the training is successful or not, it is necessary to observe the accuracy of the model's processing and judgment of unknown data. Applications increasingly operate in dynamically changing deployment scenarios, requiring optimization for both accuracy and latency [17]. The average testing accuracy of this model is about 68%, higher than the accuracy of Resnet_50 (33%) and the accuracy of Inception (30%). And the 95-percetile accuracy of FaceRecognition, Resnet_50 and Inception are 0.29, 0.18 and 0.2 respectively, while the average latency of FaceRecognition, Resnet_50 and Inception are about 0.02, 0.05 and 0.04 respectively. According to the information of the models, the accuracy of FaceRecognition is higher than other two models and it also has a lower latency, which shows that on the tasks like recognizing the people's faces, the FaceRecognition model can perform better than Resnet_50 and Inception. But the problem is that we the model was being trained, the training set and the validation set did not include data without people's faces, which results in that whenever that system scans some other objects but not faces, the model does not have the capability to distinguish the class of the object. Oppositely, other two models will identify different categories of the objects without people's faces and show the class.

FaceRecognition model has a higher throughput than other two models. The throughput can be calculated by the total requests number dived by the latency. The throughput of FaceRecognition is about 16934.53, and the throughput of Resnet_50 and Inception are about 2000 and 2500 respectively. It shows that, if there is a task which needs to recognize the people's faces in a very short time, the model Facerecognition is the best choice, but if the task requests the model to clarify the classes of different objects, the Resnet_50 and Inception should be used. The comparation of models' 95-percetile accuracy is shown in figure 2.

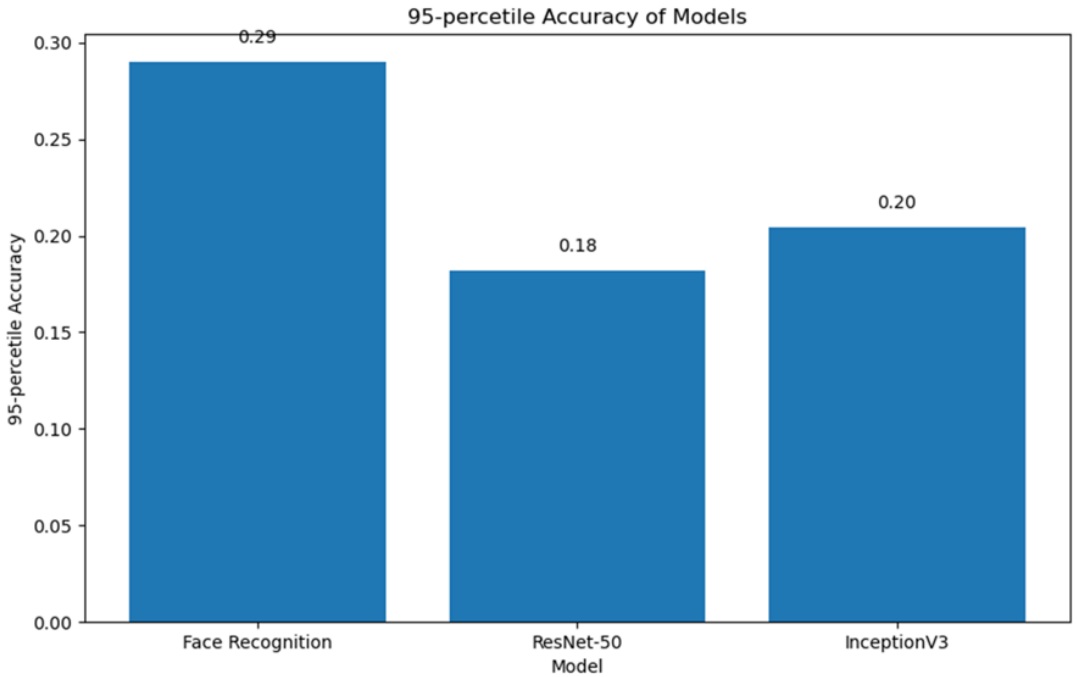

Fig. 2 The comparation of models' 95-percetile accuracy (Photo/Picture credit: Original)

The comparation of models' average accuracy is shown in figure 3.

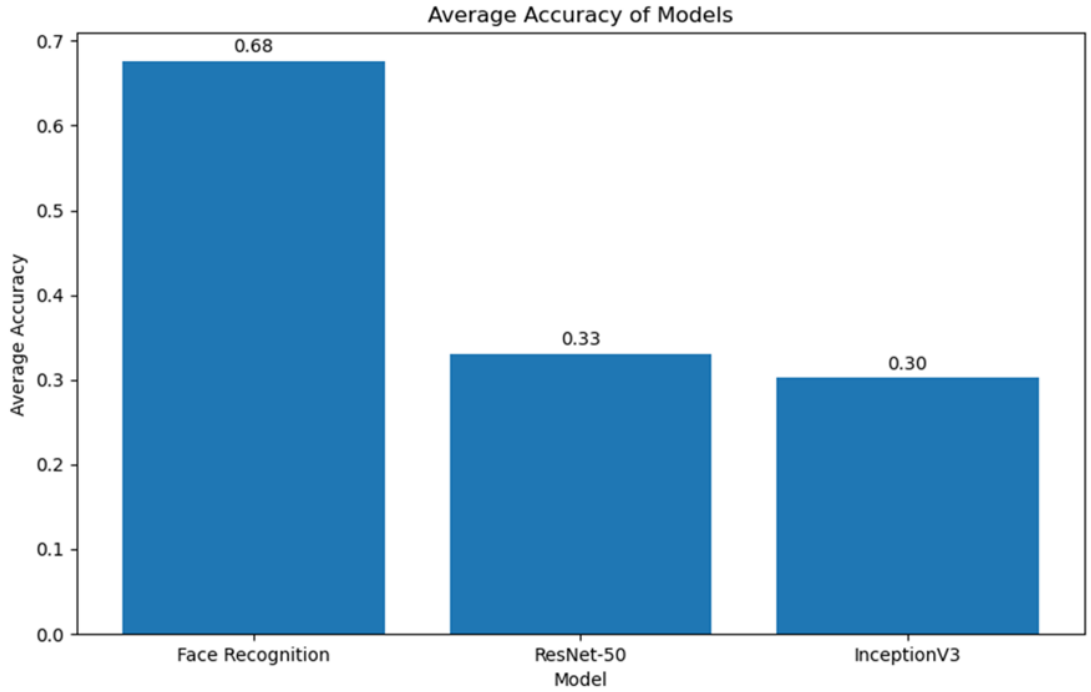

Fig. 3 The comparation of models' average accuracy (Photo/Picture credit: Original)

The comparation of models' 95-percetile latency is shown in figure 4.

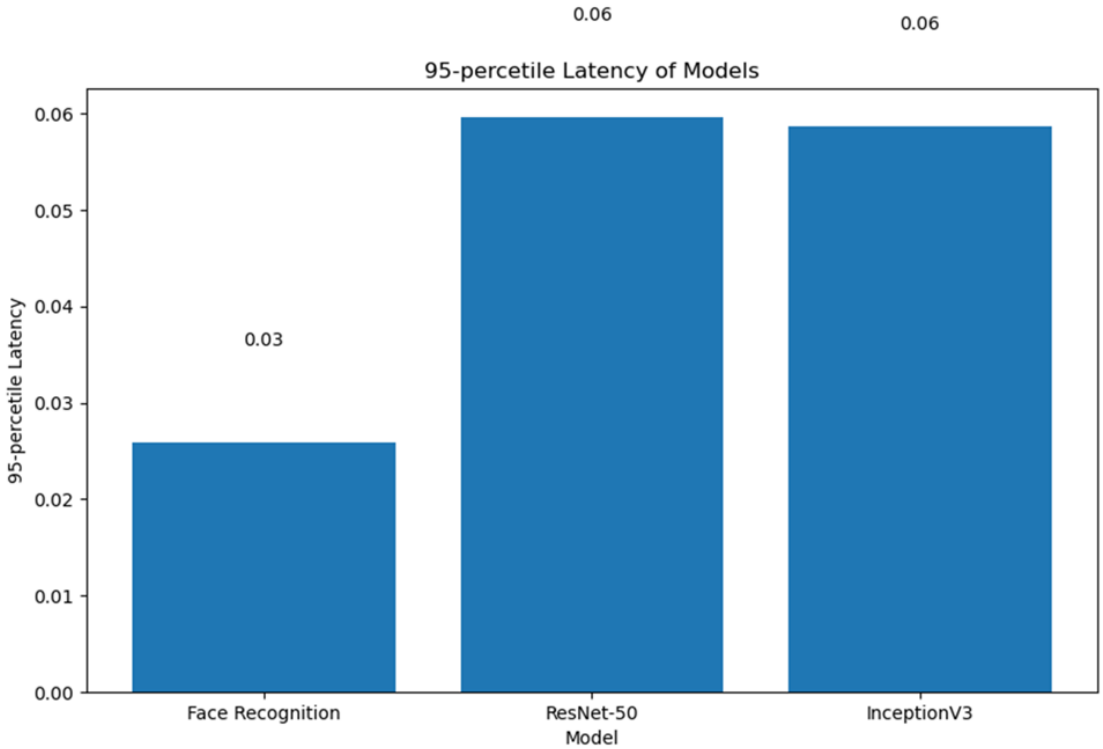

Fig. 4 The comparation of models' 95-percetile latency

The comparation of models' average latency is shown in figure 5.

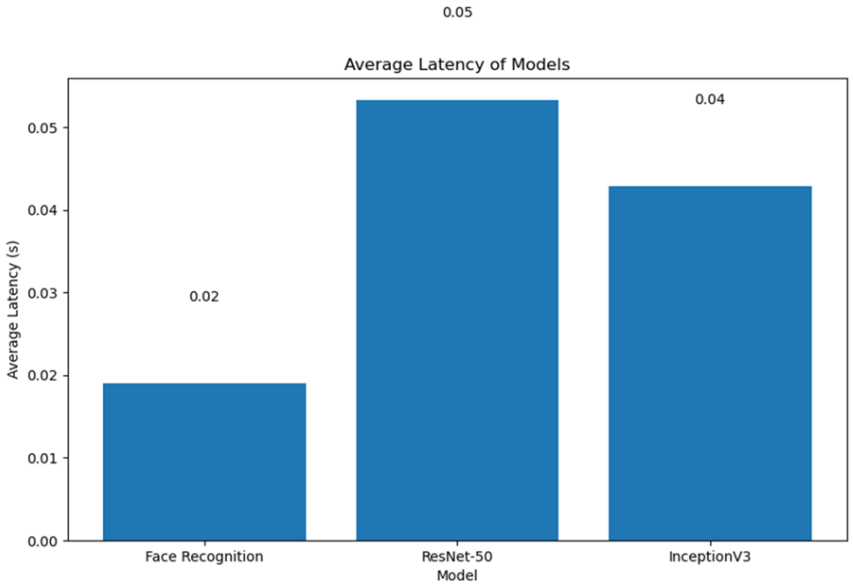

Fig. 5 The comparation of models' average latency (Photo/Picture credit: Original)

In summary, the model FaceRecognition has higher accuracy, throughput and lower latency than Resnet_50 and Inception, but it cannot distinguish the class of the object and it can only recognize the people's faces. This kind of model can achieve the numerous simple tasks in a short time, while it will not be suitable for completing the complex tasks. This also provides the opinion that to deal with elaborate situation, one of the methods is reducing the complexity of the models to improve the accuracy of achieving the simple tasks, meanwhile, increasing the number of the models so that the system can deal with different kinds of tasks. Deep learning (DL) techniques excel in capturing intricate data representations, yet their efficacy may be compromised when applied to limited datasets due to the abundance of parameters and the susceptibility of optimization algorithms to data availability, potentially resulting in overfitting. In contrast, machine learning (ML) approaches, with their relatively lower parameter counts, may exhibit greater adaptability and robustness when confronted with smaller datasets. While DL methodologies have demonstrated remarkable performance in various domains, their suitability for tasks involving classification with constrained datasets remains questionable [18].

4. Conclusion

In conclusion, the performance of FaceRecognition is different from the performance of Resnet_50 and Inception. For the FaceRecognition, it has higher accuracy, low latency and higher throughput when it is completing the face-recognition task, while the other two models have lower performance than it but can clarify the classes, and FaceRecognition cannot. According to their different performance, choosing different models to deal with different categories of situations to improve the accuracy and the efficiency is possible for humans.

The research also gives an opinion that, to improve the performance of the whole system, one method is to decrease the complexity of the sub-models and at the same time increase the number of different sub-models so that it is possible to achieve several tasks with different "professional" models. And there will be a parent model controlling sub-models. In this way, what the most important problem will be is that the performance of parent model, because if the parent model chooses the wrong sub-model to deal with the task at the beginning, the whole task cannot be completed anymore.

Actually, this kind of algorithm which is called "decision tree" has been applied in the models Resnet_50 and Inception so that these two models can distinguish the different kinds of classes of the objects and it is the commonest method to design a model which can achieve various effects. But the opinion of this article is different with that, which is willing to design a specific model for decision-making and make a better influence to human society in the future.

References

[1] Jordan I, Mitchell T M. Machine learning: Trends, perspectives, and prospects. Science, 2015, 349(6245): 255-260.

[2] Acemoglu D, Restrepo P. Artificial intelligence, automation and work. Cambridge, MA: National Bureau of Economic Research, 2018.

[3] Bostrom N, Yudkowsky E. The ethics of artificial intelligence. Mnihil P, ed. The Cambridge Handbook of Artificial Intelligence. Cambridge: Cambridge University Press, 2014: 316-334.

[4] Feigenbaum J, McCorduck P. The Fifth Generation: Artificial Intelligence and Japan's Computer Challenge to the World. Reading, MA: Addison-Wesley, 1983.

[5] Steffen W, Richardson K, Rockström J, et al. Planetary boundaries: Guiding human development on a changing planet. Science, 2015, 347(6223): 1259855-1259855.

[6] Geetha T V, Sendhilkumar S. Machine learning: concepts, techniques and applications. Boca Raton, FL: CRC Press, 2023.

[7] Paramasivam K, Sindha M M, Balakrishnan S B. KNN-Based Machine Learning Classifier Used on Deep Learned Spatial Motion Features for Human Action Recognition. Entropy, 2023, 25(6): 844. doi: 10.3390/e25060844.

[8] Glackin C E W, Adivar M. Using the Power of Machine Learning in Sales Research: Process and Potential. Journal of Personal Selling & Sales Management, 2023, 43(3): 178-194.

[9] Sundberg L, Holmström J. Using No-Code AI to Teach Machine Learning in Higher Education. Journal of Information Systems Education, 2024, 35(1): 56-66.

[10] Nežerka V, Zbíral T, Trejbal J. Machine-learning-assisted classification of construction and demolition waste fragments using computer vision: Convolution versus extraction of selected features. Expert Systems with Applications, 2024.

[11] Zeng D, Liao M, Tavakolian M, et al. Deep learning for scene classification: A survey. arxiv preprint arxiv:2101.10531, 2021.

[12] Kundu T K, Anguraj D K, Bhattacharyya D. Utilizing Machine Learning and Deep Learning for Malaria Parasite Identification in Conventional Microscopic Blood Smear Images. Traitement du Signal, 2024, 41(1): 343-362. doi: 10.18280/ts.410129.

[13] Gürbüz Y Z, Sener O, Alatan A A. Generalized sum pooling for metric learning.Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 5462-5473.

[14] Ripley B D. Pattern Recognition and Neural Networks. Cambridge: Cambridge University Press, 1996, 2007.

[15] Valliappan A M, Narayanan T N T, Kaura G. Probabilistic Loss Function Based Machine Learning System: U.S. Patent Application 17/576,851. 2023-7-20.

[16] Rao N S V. Study of Overfitting by Machine Learning Methods Using Generalization Equations. 2023 26th International Conference on Information Fusion (FUSION). 2023: 1-8.

[17] Behnam P, et al. Hardware–Software Co-Design for Real-Time Latency–Accuracy Navigation in Tiny Machine Learning Applications. IEEE Micro, 2023, 43(6): 93–101.

[18] Mirzaeian R, Nopour R, Asghari Varzaneh Z, et al. Which Machine Learning Approach is Optimal for Predicting Successful Aging: Bagging, Boosting, or Simple Algorithms?. BioMedical Engineering Online, 2023, 22(85): 1-12.